

## The Limits and Possibilities of International Large-Scale Assessments

David J. Rutkowski and Ellen L. Prusinski

VOLUME 9, NUMBER 2, SPRING 2011

### CONTENTS

<b>Introduction.....</b>	<b>1</b>
<b>Who designs and administers these assessments? .....</b>	<b>1</b>
<b>What do these assessments measure? .....</b>	<b>2</b>
<b>Who participates in the assessments and how are they organized? .....</b>	<b>2</b>
<b>What do I need to be aware of when I look at assessment results? .....</b>	<b>3</b>
<b>Why should the U.S. participate in these assessments? .....</b>	<b>3</b>
<b>Conclusion and Additional Resources .....</b>	<b>4</b>
<b>References.....</b>	<b>4</b>

### INTRODUCTION

International large-scale assessments have a long history of influencing educational policy around the world. In the United States, these assessments are often cited by policymakers and media commentators as indicators of how the American educational system is falling behind its international peers. Following the release of the 2009 Program for International Student Assessment (PISA) results, for example, Secretary of Education Arne Duncan argued, “PISA results, to be brutally honest, show that a host of developed nations are out-educating us” (Duncan, 2010). Although international assessments are certainly intended to help countries understand how their education systems compare with those of their peers, they are also intended to provide far more information than a ranking.

The staff of the Center for Evaluation & Education Policy (CEEP) at Indiana University is often asked about how international large-scale assessments influence U.S. educational policy. This policy brief is designed to provide answers to some of the most frequently asked questions encountered by CEEP researchers concerning the three most popular international education large-scale assessments: the Programme for International Student Assessment (PISA); the Trends in International Mathematics and Science Study (TIMSS); and the Progress in International Reading Literacy Study (PIRLS). These questions include:

- Who designs and administers these assessments?
- What do these assessments measure?
- Who participates in the assessments and how are they organized?
- What do I need to be aware of when I look at assessment results?
- Why should the U.S. participate in these assessments?

### UPCOMING POLICY BRIEFS ...

- ✓ *School Choice Issues in Indiana*
- ✓ *Revamping the Teacher Evaluation Process*
- ✓ *An Update on Childhood Obesity Trends*

This brief will focus, in particular, on how these three international assessments function and how they are relevant for American education. Finally, for those interested in learning more about international large-scale assessments, this brief also offers suggestions for further reading.

### WHO DESIGNS AND ADMINISTERS THESE ASSESSMENTS?

PISA is coordinated by the Organization for Economic Cooperation and Development (OECD), while TIMSS and PIRLS are both administered by the International Association for the Evaluation of Educational Achievement (IEA). Participating countries work together with these organizations to design the assessments. Each country is then responsible for administering the assessment and results are sent to the organization for data verification, processing, and scoring.

### OECD

The mission of the OECD is to promote policies that will improve the economic and social well-being of people around the world. Established in 1961, the OECD is headquartered in Paris, France, and has 34 member countries that represent the world’s most industrialized nations. The OECD analyzes and compares data, sets international standards, and recommends policies to governments around the globe.

### IEA

The IEA is an independent, international cooperative of national research institutions and government research agencies that aims to provide high-quality data capable of increasing policymakers’ understanding of key factors influencing teaching and learning. Since its founding in 1958, the IEA has conducted more than 23 research studies on

cross-national achievement. The organization is headquartered in the Netherlands with study centers in Germany and the United States. The IEA is grounded in the belief that the diversity of educational philosophies, models, and approaches that characterize the world's education systems constitute a natural laboratory in which countries can learn from one another.

## WHAT DO THESE ASSESSMENTS MEASURE?

TIMSS, PISA, and PIRLS are similar in many ways; however, they each collect unique information from different populations (see Table 1). One important distinction between IEA (TIMSS & PIRLS) and OECD (PISA) studies is that IEA studies are grade based. On the other hand, PISA differs from the TIMSS and PIRLS approach in that it samples 15-year-olds (age based), regardless of how many years of schooling students have received. Each approach has its own advantages. For example, a grade-based sample ensures that all students who take the assessment have had a similar amount of schooling, while an age-based sample can better focus on all skills attained through the first 15 years of life. As a result of sampling and other differences, the assessments should not be viewed as interchangeable and the data provided by each assessment ought to be examined separately.

### TIMSS

TIMSS provides data about trends in mathematics and science achievement of students in the fourth and eighth grades. The content assessed in TIMSS is based on an internationally agreed upon common curriculum in math and science. TIMSS collects detailed information not only about student achievement in math and science, but also about teacher preparation, resource availability, and the use of technology.

### PISA

PISA is an assessment of 15-year-old students that tests content knowledge but is not limited to school-based curricula. Instead, PISA assesses applied knowledge and literacy and emphasizes assessment of the functional skills students acquired during their schooling. The guiding question asked by PISA is "How well can students nearing the end of compulsory schooling apply their knowledge to real-life situations?" The three subject areas tested on PISA are reading literacy, mathematics literacy, and science literacy, but PISA also includes measures of general competencies such as learning strategies.

### PIRLS

PIRLS collects data to provide information on trends in reading literacy achievement of fourth-grade students. PIRLS includes an array of questions that investigate the experiences young children have at home and in school in learning to read. The assessment is offered to fourth-grade students because fourth-grade represents an important transition point in children's development as readers. In many countries, children are expected to have learned to read by fourth grade and are beginning to transition from learning to read to reading to learn. Because new countries participate in PIRLS each cycle, PIRLS also provides baseline data for new participating countries. In addition, PIRLS collects an array of information about the reading curriculum and instruction in each participating country.

it is not only countries, but also jurisdictions within countries that participate in these assessments. Thus, the U.S. is compared not only with other countries, but also provinces, cities, and linguistic communities. For example, in 2009, Shanghai, China, participated for the first time in PISA. However, to date, China has not fully participated in any of these three assessments. Therefore, care should be taken when comparing one city in China to an entire national educational system like the United States.

PIRLS, PISA, and TIMSS are administered to a sample of students so that the results can be generalized to the larger population. At the same time, it is important to recognize that each assessment defines the population to which it is generalizing (and thus from which the sample is drawn) differently.

- TIMSS: Grades 4 and 8 students (Grade-based sample)
- PIRLS: Grade 4 students (Grade-based sample)
- PISA: 15-year-old students (Age-based sample)

To ensure that the samples selected by each country are representative of the population as a whole, samples for each country are verified by an international sampling referee.

Although the specific development process for each assessment is different, each assessment is developed with concern for cross-national comparability and validity. Questions for each assessment are developed through a collaborative, international process that involves international subject area experts, national country representatives, and specialized translators. Before the assessment is administered, a field test is conducted to ensure that the assessment has low bias. The following is a brief explanation of how each assessment is administered.

**TIMSS:** Administered every four years, TIMSS began in 1995. The next round of TIMSS is currently taking place (in 2011), with over 60 participating systems. The data

## WHO PARTICIPATES IN THE ASSESSMENTS AND HOW ARE THEY ORGANIZED?

There is great diversity among the countries and economies that participate in the assessments, including diversity in economic development, geographical location, language, and size. In fact, it is important to recognize that

TABLE 1. Overview of PISA, TIMSS, AND PIRLS

	PISA	TIMSS	PIRLS
How often are tests conducted?	Every 3 years	Every 4 years	Every 5 years
Who is tested?	15-year-old students	4th and 8th grade students	4th grade students
What is tested?	Ability to apply skills and competencies to "real-world" contexts	Math and Science curricula	Reading curriculum
Who sponsors the test?	OECD	IEA	IEA
How many systems participated in the last cycle?	65	60	55

are scheduled to be released in the end of 2012.

**PISA:** PISA was first administered in 2000 and is offered every three years. In 2009, the most recent round of testing, 65 countries and other education systems—including all 34 OECD countries, 26 non-OECD countries, and 5 non-national education systems—participated in PISA. The data for the next round of PISA will be made available in 2012.

**PIRLS:** PIRLS was first administered in 2001 and is offered every five years. The next round of PIRLS is currently taking place (in 2011), with over 55 participating systems. The data are scheduled to be released in the end of 2012.

### WHAT DO I NEED TO BE AWARE OF WHEN I LOOK AT ASSESSMENT RESULTS?

TIMSS, PISA, and PIRLS all provide data that can be used to compare countries to one another. Although each country has an aggregate achievement score on the assessment, the ranking of countries does not always indicate significant differences in achievement. For example, on the PISA 2009 assessment, the U.S. average score of 500 was not significantly different from the OECD average score of 493. Although the U.S. was ranked 14th overall on reading literacy on the PISA 2009, only six OECD countries, including Korea, Finland, and Canada, had significantly higher average scores (see Table 2). After accounting for standard errors, 14 OECD countries, including Germany, Norway, and Poland, had results that were not measurably different from the U.S. Thus, although assessments results are comparable cross-nationally, it is difficult to specifically rank countries. Rather, scores provide countries with information on where they generally rank compared to others.

Additionally, because of the complexity of the study design, scores cannot be compared at the individual level. In other words, it is never appropriate to compare one student to another student using these assessments. One reason this is not possible is because of the volume of information covered by each assessment. Because each assessment aims to cover a large amount of information, no student takes the entire test. This does not mean that results are not valid or reliable at aggregated levels. Rather, advanced and proven statistical techniques are used to infer five possible scores for each participant, which

**TABLE 2. PISA Reading Scores**

Combined Reading Literacy Scale	
OECD Average	493
<i>OECD Countries</i>	
Korea, Republic of	539
Finland	536
Canada	524
New Zealand	521
Japan	520
Australia	515
Netherlands	508
Belgium	506
Norway	503
Estonia	501
Switzerland	501
Poland	500
Iceland	500
United States	500
Sweden	497
Germany	497
Ireland	496
France	496
Denmark	495
United Kingdom	494
Hungary	494
Portugal	489
Italy	486
Slovenia	483
Greece	483
Spain	481
Czech Republic	478
Slovak Republic	477
Israel	474
Luxembourg	472
Austria	470
Turkey	454
Chile	449
Mexico	425

	Average is higher than the U.S. average
	Average is not measurably different from the U.S. average
	Average is lower than the U.S. average

are then used to create scores for populations. Analysis of the resulting data is possible but requires sophisticated analysis techniques.

Finally, when looking at international assessment results, it is vital to remember that populations change. Although the assessments allow for consideration of changes in a country's results over time, changing demographic patterns and country contexts mean that changes in scores must be interpreted with caution. When looking at scores over time, the country context, education system, and demographic composition must be kept in mind.

### WHY SHOULD THE U.S. PARTICIPATE IN THESE ASSESSMENTS?

As described above, there are limits to what the assessments are able to show us about the American educational system. However, participation in the assessments also offers valuable information capable of pointing out trends in education and student achievement. For example, participation in TIMSS, PISA, and PIRLS can help countries:

- determine their global educational standing in subjects essential for further learning, including reading, mathematics, and science,
- profile relative strengths and weaknesses in reading, mathematics, and science achievement in an international context,
- measure educational progress over time both within and between countries,
- inform national and local policy about schools' curricula and instruction,
- collect in-depth information about school environments, resources, and instruction, and
- examine concerns about equity in learning opportunities.

Another common question related to U.S. participation in international assessment is why the U.S. should participate when American students also regularly participate in the National Assessment of Educational Progress (NAEP). Although NAEP is similar to TIMSS, PISA, and PIRLS in that it measures students' performance in reading, mathematics, and science, it generally cannot be used to benchmark the U.S. performance to that of other countries because NAEP is designed specifically to meet national and state information needs.

More broadly, participation in international large-scale assessments can help the U.S.

think comparatively about its education system. Although the American education system is certainly unique, it may not be as different from other countries as is often assumed. For example, in the TIMSS 2007 fourth-grade sample, Algeria, Australia, Hong Kong, and New Zealand, among other countries, all had percentages of immigrants that were greater than the U.S.

Additionally, Williams (2003), points out that thinking comparatively about education can help us learn from others, help us understand and work with others, and help us understand ourselves. As Williams describes, “International comparison provides surprising insights” into the status of the American education system and helps us better understand ourselves. Among the most striking examples of such insights is the breakdown of the 2009 PISA results by racial/ethnic groups in the U.S. When broken out by race/ethnicity, Asian-American students ranked among the highest achieving students in the world. Additionally, White American students had scores similar to the top-performing countries, New Zealand and Finland. However, Hispanic American and Black American participants fell far below the international average. In fact, these groups of students are grouped with the lowest OECD members of Turkey, Chile, and Mexico. Such results raise troubling questions about the persistent inequality in American education and suggest that more needs to be done to ensure that all American students, regardless of race/ethnicity or socio-economic class, receive a high-quality education.

## CONCLUSION AND ADDITIONAL RESOURCES

International assessments such as PISA, TIMSS, and PIRLS provide valuable information on the American education system. Because each international assessment is unique in its goals, format, and content, results of each assessment must be understood as providing distinctive information that, when taken together, contribute to a fuller understanding of how the American education system compares with education systems in other countries. Because no one educational system dominates the design of these assessments, these three international assessments differ in many ways from national assessments such as the National Assessment of Educational Progress (NAEP). To allow specific states to compare to other national educational systems, the National Center for Education Statistics is developing a new study to link national and international student assessments so that states can measure their performance against international benchmarks. It should be noted that linking two assessments such as NAEP and TIMSS is a complicated process and interpretations resulting from the study should be taken with caution.

Additionally, because the cross-country comparisons based on international assessment data found in the popular media often misuse or oversimplify the results of these assessments, it is critical to understand how the tests are developed, administered, and analyzed. In order to help readers better understand the reports they encounter in the media and the ways policymakers use international assessment results in their policy arguments, this brief has offered an overview of the tests

and provided answers to some of the most commonly asked questions regarding international assessment.

If you are interested in learning more about the organizations who administer these assessments or how TIMSS, PISA, and PIRLS are developed, administered, and analyzed, the following resources offer additional information about the assessments:

**OECD**  
[www.oecd.org](http://www.oecd.org)

**IEA**  
[www.iea.nl](http://www.iea.nl)  
**TIMSS**  
[timss.bc.edu/timss2011/index.html](http://timss.bc.edu/timss2011/index.html)

**PISA**  
[www.pisa.oecd.org](http://www.pisa.oecd.org)

**PIRLS**  
[pirls.bc.edu/pirls2011/index.html](http://pirls.bc.edu/pirls2011/index.html)

**NCES International Activities**  
[nces.ed.gov/surveys/international/](http://nces.ed.gov/surveys/international/)

**NAEP-TIMSS**  
<http://nces.ed.gov/timss/naeplink.asp>

**NAEP**  
<http://nces.ed.gov/nationsreportcard/>

## References

- Duncan, A. (2010, December). Secretary Arne Duncan's remarks at OECD's release of the Program for International Student Assessment (PISA) 2009 Results. Available at: <http://www.ed.gov/news/speeches/secretary-arne-duncans-remarks-oecd-s-release-program-international-student-assessment>
- Williams, J.H. (2003). Why compare? Why all educators should think internationally. *International Educator*, 18-25.

Education Policy Briefs are executive edited by Jonathan A. Plucker, Ph.D. and published by the

## Center for Evaluation & Education Policy

Indiana University  
1900 East Tenth Street  
Bloomington, IN 47406-7512  
812-855-4438

More about the Center for Evaluation & Education Policy  
and our publications can be found at our Web site:  
<http://ceep.indiana.edu>



CENTER FOR EVALUATION  
& EDUCATION POLICY

0801110106

INDIANA UNIVERSITY  
SCHOOL OF EDUCATION